

# Project Data Analysis for B2B Retail: Customer Analytics Report

## Latar Belakang

xyz.com adalah perusahaan rintisan B2B yang menjual berbagai produk tidak langsung kepada end user tetapi ke bisnis/perusahaan lainnya. Sebagai data-driven company, maka setiap pengambilan keputusan di xyz.com selalu berdasarkan data. Setiap quarter xyz.com akan mengadakan townhall dimana seluruh atau perwakilan divisi akan berkumpul untuk me-review performance perusahaan selama quarter terakhir.

## Tugas dan Langkah

Menyediakan data dan analisa mengenai kondisi perusahaan bulan terakhir untuk dipresentasikan di townhall tersebut. (Asumsi tahun yang sedang berjalan adalah tahun 2004).

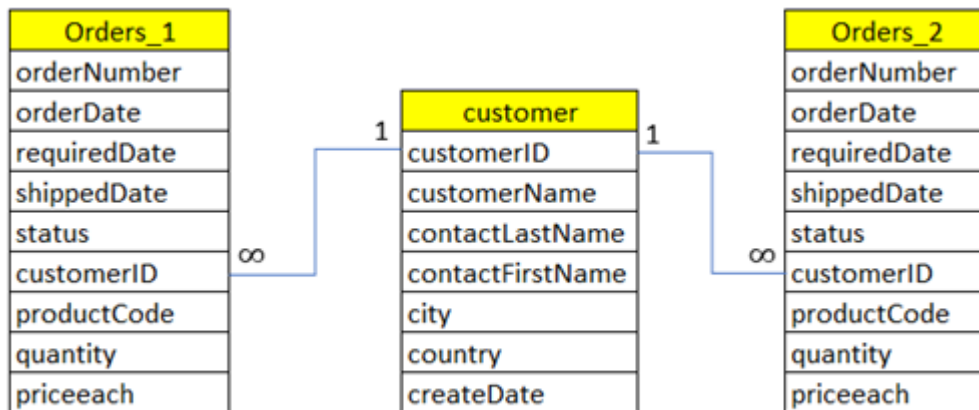
Adapun hal yang akan direview adalah :

1. Bagaimana pertumbuhan penjualan saat ini?
2. Apakah jumlah customers xyz.com semakin bertambah ?
3. Dan seberapa banyak customers tersebut yang sudah melakukan transaksi?
4. Category produk apa saja yang paling banyak dibeli oleh customers?
5. Seberapa banyak customers yang tetap aktif bertransaksi?

## Tabel yang Digunakan

Tabel yang akan digunakan pada project kali ini adalah sebagai berikut.

1. Tabel orders\_1 : Berisi data terkait transaksi penjualan periode quarter 1 (Jan – Mar 2004)
2. Tabel Orders\_2 : Berisi data terkait transaksi penjualan periode quarter 2 (Apr – Jun 2004)
3. Tabel Customer : Berisi data profil customer yang mendaftar menjadi customer xyz.com



## Memahami table

Sebelum membuat memulai menyusun query SQL dan membuat Analisa dari hasil query, hal yang perlu dilakukan adalah pertama-tama menjadi familiar dengan tabel yang akan digunakan. Hal ini akan sangat berguna dalam menentukan kolom mana sekiranya berkaitan dengan problem yang akan dianalisa, dan proses manipulasi data apa yang sekiranya perlu dilakukan untuk kolom – kolom tersebut, karena tidak semua kolom pada tabel perlu untuk digunakan.

```

1 SELECT * FROM orders_1 limit 5;
2 SELECT * FROM orders_2 limit 5;
3 SELECT * FROM customer limit 5;
  
```

**Tabel orders\_1**

orderNumber	orderDate	requiredDate	shippedDate	status	customerID	productCode	quantity	priceeach
10234	2004-03-30	2004-04-05	2004-04-02	Shipped	412	S72_1253	40	45690
10234	2004-03-30	2004-04-05	2004-04-02	Shipped	412	S700_2047	29	83280
10234	2004-03-30	2004-04-05	2004-04-02	Shipped	412	S24_3816	31	78830
10234	2004-03-30	2004-04-05	2004-04-02	Shipped	412	S24_3420	25	65090
10234	2004-03-30	2004-04-05	2004-04-02	Shipped	412	S24_2841	44	67140

**Table orders\_2**

orderNumber	orderDate	requiredDate	shippedDate	status	customerID	productCode	quantity	priceeach
10235	2004-04-02	2004-04-12	2004-04-06	Shipped	260	S18_2581	24	81950
10235	2004-04-02	2004-04-12	2004-04-06	Shipped	260	S24_1785	23	89720
10235	2004-04-02	2004-04-12	2004-04-06	Shipped	260	S24_3949	33	55270
10235	2004-04-02	2004-04-12	2004-04-06	Shipped	260	S24_4278	40	63030
10235	2004-04-02	2004-04-12	2004-04-06	Shipped	260	S32_1374	41	90900

**Table customer**

customerID	customerName	contactLastName	contactFirstName	city	country	createDate
103	Atelier graphique	Schmitt	Carine	Nantes	France	2004-02-05
112	Signal Gift Stores	King	Jean	Las Vegas	USA	2004-02-05
114	Australian Collectors, Co.	Ferguson	Peter	Melbourne	Australia	2004-02-20
119	La Rochelle Gifts	Labrune	Janine	Nantes	France	2004-02-05
121	Baane Mini Imports	Bergulfsen	Jonas	Stavern	Norway	2004-02-05

## Total Penjualan dan Revenue pada Quarter-1 (Jan, Feb, Mar) dan Quarter-2 (Apr, Mei, Jun)

1. Dari tabel orders\_1 lakukan penjumlahan pada kolom quantity dengan fungsi aggregate sum() dan beri nama "total\_penjualan", kalikan kolom quantity dengan kolom priceEach kemudian jumlahkan hasil perkalian kedua kolom tersebut dan beri nama "revenue"
2. Perusahaan hanya ingin menghitung penjualan dari produk yang terkirim saja, jadi kita perlu mem-filter kolom 'status' sehingga hanya menampilkan order dengan status "Shipped".
3. Lakukan Langkah 1 & 2, untuk tabel orders\_2.

### SQL Query :

```
1 select sum(quantity) as total_penjualan, sum(quantity * priceeach) as revenue from orders_1 where
   status='shipped';
2
3 select sum(quantity) as total_penjualan, sum(quantity * priceeach) as revenue from orders_2 where
   status='shipped';
```

### Hasilnya sebagai berikut :

total_penjualan	revenue
8694	799579310
total_penjualan	revenue
6717	607548320

# Menghitung persentasi keseluruhan penjualan

Kedua tabel orders\_1 dan orders\_2 masih terpisah, untuk menghitung persentasi keseluruhan penjualan dari kedua tabel tersebut perlu digabungkan :

1. Pilih kolom "orderNumber", "status", "quantity", "priceEach" pada tabel orders\_1, dan tambahkan kolom baru dengan nama "quarter" dan isi dengan value "1". Lakukan yang sama dengan tabel orders\_2, dan isi dengan value "2", kemudian gabungkan kedua tabel tersebut.
2. Gunakan statement dari Langkah 1 sebagai subquery dan beri alias "tabel\_a".
3. Dari "tabel\_a", lakukan penjumlahan pada kolom "quantity" dengan fungsi aggregate sum() dan beri nama "total\_penjualan", dan kalikan kolom quantity dengan kolom priceEach kemudian jumlahkan hasil perkalian kedua kolom tersebut dan beri nama "revenue"
4. Filter kolom 'status' sehingga hanya menampilkan order dengan status "Shipped".
5. Kelompokkan total\_penjualan berdasarkan kolom "quarter", dan jangan lupa menambahkan kolom ini pada bagian select.

## SQL Query :

```
1 |select quarter,sum(quantity) as total_penjualan,sum(quantity * priceEach) as
   |revenue from(select
2 |orderNumber,
3 |status,
4 |quantity,
5 |priceEach,
6 |1 as quarter
7 |from orders_1
8 |union
9 |select
10|orderNumber,
11|status,
12|quantity,
13|priceEach,
14|2 as quarter
15|from orders_2)as tabel_a where status='shipped' group by quarter;
```

## Hasil kolom sebagai berikut :

quarter	total_penjualan	revenue
1	8694	799579310
2	6717	607548320

# Perhitungan Growth Penjualan dan Revenue

Untuk project ini, perhitungan pertumbuhan penjualan akan dilakukan secara manual menggunakan formula yang disediakan diatas.

$\% \text{Growth Penjualan} = (6717 - 8694) / 8694 = -22\%$

$\% \text{Growth Revenue} = (607548320 - 799579310) / 799579310 = -24\%$

## Apakah jumlah customers xyz.com semakin bertambah?

Penambahan jumlah customers dapat diukur dengan membandingkan total jumlah customers yang registrasi di periode saat ini dengan total jumlah customers yang registrasi diakhir periode sebelumnya.

1. Dari tabel customer, pilihlah kolom customerID, createDate dan tambahkan kolom baru dengan menggunakan fungsi QUARTER(...) untuk mengekstrak nilai quarter dari CreateDate dan beri nama "quarter"
2. Filter kolom "createDate" sehingga hanya menampilkan baris dengan createDate antara 1 Januari 2004 dan 30 Juni 2004
3. Gunakan statement Langkah 1 & 2 sebagai subquery dengan alias tabel\_b
4. Hitunglah jumlah unik customers à tidak ada duplikasi customers dan beri nama "total\_customers"
5. Kelompokkan total\_customer berdasarkan kolom "quarter", dan jangan lupa menambahkan kolom ini pada bagian select.

### SQL Query :

```
1 select quarter, count(customerID) from (select
  customerID, createDate, quarter(createDate) as quarter from customer where
2 createDate between '2004-01-01' and '2004-06-30') tabel_b group by quarter;
3 |
```

### Hasil kolom sebagai berikut :

quarter	count(customerID)
1	43
2	35

# Seberapa banyak customers tersebut yang sudah melakukan transaksi?

Problem ini merupakan kelanjutan dari problem sebelumnya yaitu dari sejumlah customer yang registrasi di periode quarter-1 dan quarter-2, berapa banyak yang sudah melakukan transaksi

1. Dari tabel customer, pilih kolom customerID, createDate dan tambahkan kolom baru dengan menggunakan fungsi QUARTER(...) untuk mengekstrak nilai quarter dari CreateDate dan beri nama "quarter"
2. Filter kolom "createDate" sehingga hanya menampilkan baris dengan createDate antara 1 Januari 2004 dan 30 Juni 2004
3. Gunakan statement Langkah A&B sebagai subquery dengan alias tabel\_b
4. Dari tabel orders\_1 dan orders\_2, pilihlah kolom customerID, gunakan DISTINCT untuk menghilangkan duplikasi, kemudian gabungkan dengan kedua tabel tersebut dengan UNION.
5. Filter tabel\_b menggunakan operator IN() dan 'Select statement Langkah d', sehingga hanya customerID yang pernah bertransaksi (customerID tercatat di tabel orders) yang diperhitungkan.
6. Hitunglah jumlah unik customers (tidak ada duplikasi customers) di statement SELECT dan beri nama "total\_customers"
7. Kelompokkan total\_customer berdasarkan kolom "quarter", dan jangan lupa menambahkan kolom ini pada bagian select.

## SQL Query :

```
1 select quarter, count(distinct customerID) as total_customers from (select
customerID, createDate, quarter(createDate) as quarter from customer where createDate
between '2004-01-01' and '2004-06-30') as tabel_b where customerID in (select
distinct customerID from orders_1
2 union
3 select distinct customerID from orders_2) group by quarter;
```

## Hasil kolom sebagai berikut :

quarter	total_customers
1	25
2	19

# Category produk apa saja yang paling banyak di-order oleh customers di Quarter-2?

Untuk mengetahui kategori produk yang paling banyak dibeli, maka dapat dilakukan dengan menghitung total order dan jumlah penjualan dari setiap kategori produk.

1. Dari kolom orders\_2, pilih productCode, orderNumber, quantity, status
2. Tambahkan kolom baru dengan mengekstrak 3 karakter awal dari productCode yang merupakan ID untuk kategori produk; dan beri nama categoryID
3. Filter kolom "status" sehingga hanya produk dengan status "Shipped" yang diperhitungkan
4. Gunakan statement Langkah 1, 2, dan 3 sebagai subquery dengan alias tabel\_c
5. Hitunglah total order dari kolom "orderNumber" dan beri nama "total\_order", dan jumlah penjualan dari kolom "quantity" dan beri nama "total\_penjualan"
6. Kelompokkan berdasarkan categoryID, dan jangan lupa menambahkan kolom ini pada bagian select.
7. Urutkan berdasarkan "total\_order" dari terbesar ke terkecil.

## SQL Query

```
1 SELECT * FROM (SELECT categoryID, COUNT(DISTINCT orderNumber) AS
total_order, SUM(quantity) AS total_penjualan
2 FROM (
3 SELECT
4 productCode, |
5 orderNumber,
6 quantity,
7 status,
8 LEFT(productCode,3) AS categoryID
9 FROM orders_2
10 WHERE status = "Shipped") tabel_c
11 GROUP BY categoryID ) a
12 ORDER BY total_order DESC
```

Hasil kolom sebagai berikut :

categoryID	total_order	total_penjualan
S18	25	2264
S24	21	1826
S32	11	616
S12	10	491
S50	8	292
S10	8	492
S70	7	675
S72	2	61

## Seberapa banyak customers yang tetap aktif bertransaksi setelah transaksi pertamanya?

Mengetahui seberapa banyak customers yang tetap aktif menunjukkan apakah xyz.com tetap digemari oleh customers untuk memesan kebutuhan bisnis mereka. Hal ini juga dapat menjadi dasar bagi tim product dan business untuk pengembangan product dan business kedepannya. Adapun metrik yang digunakan disebut retention cohort. Untuk project ini, kita akan menghitung retention dengan query SQL sederhana.

Oleh karena baru terdapat 2 periode yang Quarter 1 dan Quarter 2, maka retention yang dapat dihitung adalah retention dari customers yang berbelanja di Quarter 1 dan kembali berbelanja di Quarter 2, sedangkan untuk customers yang berbelanja di Quarter 2 baru bisa dihitung retentionnya di Quarter 3.

1. Dari tabel orders\_1, tambahkan kolom baru dengan value "1" dan beri nama "quarter"
2. Dari tabel orders\_2, pilihlah kolom customerID, gunakan distinct untuk menghilangkan duplikasi
3. Filter orders\_1 menggunakan operator IN() dan 'Select statement Langkah 2)', sehingga hanya customerID yang pernah bertransaksi di quarter 2 (customerID tercatat di tabel orders\_2) yang diperhitungkan.
4. Hitunglah jumlah unik customers (tidak ada duplikasi customers) dibagi dengan total\_ customers dalam percentage, pada Select statement dan beri nama "Q2"



### SQL Query :

```
1 SELECT
2 COUNT(DISTINCT customerID) as total_customers
3 FROM orders_1 ;
4 SELECT
5     '1' as quarter,
6     (COUNT(DISTINCT customerID)/25)*100 as Q2
7 FROM
8     orders_1
9 WHERE customerID IN(
10     SELECT
11         distinct customerID
12     FROM
13         orders_2)
```

### Hasil kolom sebagai berikut :

```
+-----+
| total_customers |
+-----+
|                25 |
+-----+
+-----+-----+
| quarter | Q2      |
+-----+-----+
| 1       | 24.0000 |
+-----+-----+
```

## Kesimpulan

Berdasarkan data yang telah kita peroleh melalui query SQL, Kita dapat menarik kesimpulan bahwa :

1. Performance xyz.com menurun signifikan di quarter ke-2, terlihat dari nilai penjualan dan revenue yang drop hingga 20% dan 24%,
2. perolehan customer baru juga tidak terlalu baik, dan sedikit menurun dibandingkan quarter sebelumnya.
3. Ketertarikan customer baru untuk berbelanja di xyz.com masih kurang, hanya sekitar 56% saja yang sudah bertransaksi. Disarankan tim Produk untuk perlu mempelajari behaviour customer dan melakukan product improvement, sehingga conversion rate (register to transaction) dapat meningkat.
4. Produk kategori S18 dan S24 berkontribusi sekitar 50% dari total order dan 60% dari total penjualan, sehingga xyz.com sebaiknya fokus untuk pengembangan category S18 dan S24.

5. Retention rate customer xyz.com juga sangat rendah yaitu hanya 24%, artinya banyak customer yang sudah bertransaksi di quarter-1 tidak kembali melakukan order di quarter ke-2 (no repeat order).
6. XYZ.com mengalami pertumbuhan negatif di quarter ke-2 dan perlu melakukan banyak improvement baik itu di sisi produk dan bisnis marketing, jika ingin mencapai target dan positif growth di quarter ke-3. Rendahnya retention rate dan conversion rate bisa menjadi diagnosa awal bahwa customer tidak tertarik/kurang puas/kecewa berbelanja di xyz.com.

## Data Mentor

*Trisna Yulia Junita*

Data Scientist PT. BUMA